

**IN THE CLAIMS:**

Please cancel claims 1-14.

15. (new) A method of discretization/grouping of a source attribute or a source attributes group of a database containing a population of individuals with the object in particular of predicting modalities of a given target attribute, said method comprising the following steps of:

- (a) partitioning of said modalities of said source attribute or said attributes group into elementary regions,
- (b) evaluating of a merge criterion for each pair of elementary regions,
- (c) searching, among the set of pairs of elementary regions that can be merged, for the pair of elementary regions for which the merge criterion would be optimized,
- (d) skipping to step f) as long as the value of a valuation variable of the merge under consideration, said valuation variable characterizing the behavior of said merge criterion, is not within a predetermined zone of atypical values,
- (e) stopping the method if there are no elementary regions whose merge would have a consequence of improving said merge criterion, and
- (f) otherwise merging and reiterating of steps b) to e).

16. (new) A method of discretization/grouping of a source attribute or source attributes group according to claim 1, wherein said predetermined zone of atypical values is such that for a target attribute independent of said source attribute or said source attributes group, the value of said valuation variable of the merge under consideration is not within said zone with a predetermined probability  $p$ .

17. (new) A method of discretization of a source attribute of a database containing a population of individuals with the object in particular of predicting modalities of a given target attribute, said method comprising the following steps of:

- (a) partitioning of said modalities of the source attribute into adjacent two-by-two elementary intervals.
- (b) evaluating for each pair of adjacent elementary intervals of said set of the value of  $\chi^2$  of a contingency table after a possible merge of said pair,
- (c) searching, among the set of pairs of elementary intervals that can be merged, for the pair of elementary intervals whose merge would maximize the value of  $\chi^2$ ,
- (d) skipping directly to step f) as long as the value  $\Delta\chi^2$  of the variation of the value of  $\chi^2$  before and after merge is, in absolute value, less than a predetermined threshold value  $\text{Max}\Delta\chi^2$ ,
- (e) stopping of the method if there are no elementary intervals that make it possible to reduce a probability of independence, and
- (f) otherwise merging and reiterating of steps b) to e).

18. (new) A discretization method according to claim 17, wherein said predetermined threshold value  $\text{Max}\Delta\chi^2$  is such that for a target attribute independent of the source attribute the value  $\Delta\chi^2$  of the variation of the value of  $\chi^2$  before and after merge is always less than said value  $\text{Max}\Delta\chi^2$  with a predetermined probability p.

19. (new) A discretization method according to claim 18, wherein said predetermined threshold value  $\text{Max}\Delta\chi^2$  is equal to the function of  $\chi^2$  of degree of freedom equal to the number J of modalities of the target attribute minus one for a second probability p to the power  $1/N$  where N is the size of the sample of the part of the database to which said discretization method is applied:

$$\text{Max}\Delta\chi^2 = \text{Inv}\chi^2_{J-1}(p^{1/N})$$

where  $Inv\chi^2$  is the function that gives the value of  $\chi^2$  as a function of a given probability p.

20. (new) A method of discretization of a source attribute according to claim 19, further comprising a step of verification that the effectiveness of the source attribute for modalities in a given interval for each target attribute is greater than the predetermined value, and if such is not the case, to implement the merge of said interval with an adjacent interval.

21. (new) A method of grouping of a source attribute of a database containing a population of individuals with the object in particular of predicting modalities of a given target attribute, said method comprising the following steps of:

- (a) partitioning of said modalities of the source attribute into a plurality of groups,
- (b) evaluating for each plurality of groups of said set of the value of  $\chi^2$  of a contingency table after a possible merge of said plurality of groups,
- (c) searching among the set of plurality of groups that can be merged for the groups whose merge would maximize the value of  $\chi^2$ ,
- (d) skipping directly to step f) as long as the value  $\Delta\chi^2$  of the variation of the value of  $\chi^2$  before and after merge is, in absolute value, less than a predetermined threshold value  $\text{Max}\Delta\chi^2$ ,
- (e) stopping of the method if there are no merges of groups that make it possible to reduce a probability of independence, and
- (f) otherwise merging and reiteration of steps b) to e).

22. (new) A grouping method according to claim 21, wherein said predetermined threshold value  $\text{Max}\Delta\chi^2$  is such that for a target attribute independent of the source attribute the value  $\Delta\chi^2$  of the variation of the value of  $\chi^2$  before and after merge is always less than said value  $\text{Max}\Delta\chi^2$  with a predetermined probability p.

23. (new) A grouping method according to claim 22, wherein establishing the predetermined threshold value  $\text{Max}\Delta\chi^2$  consists in using a previously calculated table of values of mean and standard deviation as a function of the number of modalities of the source attribute and of the number of modalities of the target attributes to determine by linear interpolation from said table of values the mean and standard deviation of  $\text{Max}\Delta\chi^2$  corresponding to the attributes to be grouped, and then to determine, by using the inverse normal law, the corresponding predetermined threshold value  $\text{Max}\Delta\chi^2$  which will not be with the probability p.

24. (new) A grouping method according to claim 23, wherein for two target modalities, the mean of  $\text{Max}\Delta\chi^2$  is asymptotically proportional to  $2I/\pi$ , where I is the number of the source modalities.

25. (new) A grouping method according to claim 24, wherein for two source modalities, the law of  $\text{Max}\Delta\chi^2$  is the law of  $\chi^2$  with J-1 degrees of freedom, J being the number of target modalities.

26. (new) A method of grouping of a source attribute according to claim 25, further comprising a preliminary step of verifying that the effectiveness of the source attribute for modalities in a given group for each target attribute is greater than the predetermined value, and if such is not the case, to implement a merge of said group with a specific group, said merged group then forming again said specific group.

27. (new) A method of discretization in dimension  $k$  of a group of  $k$  continuous source attributes of a database containing a population of individuals, with the object in particular of predicting the modalities of a given target attribute, said method comprising the following steps of:

- (a) partitioning of said modalities of the group of  $k$  source attributes into elementary regions of dimension  $k$ ,
- (b) evaluating for elementary regions of dimension  $k$  of the value of  $\chi^2$  of a contingency table after a possible merge of said elementary regions of dimension  $k$ ,
- (c) searching among the set of said elementary regions of dimension  $k$  that can be merged, for the elementary regions of dimension  $k$  whose merge would maximize the value of  $\chi^2$ ,
- (d) skipping directly to step f) as long as the value  $\Delta\chi^2$  of the variation of the value of  $\chi^2$  before and after merge is, in absolute value, less than a predetermined threshold value  $\text{Max}\Delta\chi^2$ ,
- (e) stopping of the method if there is no set of intervals that make it possible to reduce a probability of independence, and
- (f) otherwise merging and reiterating of steps b) to e).

28. (new) A method of grouping in dimension  $k$  of a group of  $k$  discrete source attributes of a database containing a population of individuals, with the object in particular of predicting the modalities of a given target attribute, said method comprising the following steps of:

- (a) partitioning of said modalities of the group of  $k$  source attributes into a plurality of groups,
- (b) evaluating for each plurality of groups of the value of  $\chi^2$  of a contingency table after a possible merge of said plurality,
- (c) searching, among the set of plurality of groups that can be merged, for the plurality of groups whose merge would maximize the value of  $\chi^2$ ,
- (d) skipping directly to step f) as long as the value  $\Delta\chi^2$  of the variation of the value of  $\chi^2$  before and after merge is, in an absolute value, less than a predetermined threshold value  $\text{Max}\Delta\chi^2$ ,
- (e) stopping of the method if there is no set of intervals that make it possible to reduce a probability of independence, and
- (f) otherwise merging and reiterating of steps b) to e).